



# UNITED STATES PATENT AND TRADEMARK OFFICE

MN

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/833,846	04/12/2001	Edward Clifford Kubaitis	08214/1200332-US2	2227
7278	7590	05/10/2007		
DARBY & DARBY P.C. P. O. BOX 5257 NEW YORK, NY 10150-5257			EXAMINER TRUONG, CAM Y T	
			ART UNIT	PAPER NUMBER
			2162	
			MAIL DATE	DELIVERY MODE
			05/10/2007	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<b>Office Action Summary</b>	<b>Application No.</b> 09/833,846	<b>Applicant(s)</b> KUBAITIS, EDWARD CLIFFORD	
	<b>Examiner</b> Cam Y T. Truong	<b>Art Unit</b> 2162	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) ☒ Responsive to communication(s) filed on 03 May 2007.
- 2a) ☐ This action is **FINAL**.                      2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) ☒ Claim(s) 1-38 and 41-43 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-38 and 41-43 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)                                | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                       | 5) <input type="checkbox"/> Notice of Informal Patent Application                       |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

### **DETAILED ACTION**

1. Applicant has amended claims 1, 11, 17, 27 and 34 and added claims 41-43 in the amendment filed on 4/6/2007.

Claims 1-38 and 41-43 are pending in this Office Action.

### ***Response to Arguments***

2. Applicant's arguments with respect to claims 41-43 have been considered but are moot in view of the new ground(s) of rejection.

a. Applicant argued that Bates does not teach "providing extracted data from the determined web domain address in a data log directly to the user".

However, Madnick teaches returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

Thus, the combination of cited reference teaches the above claimed limitation.

b. Applicant argued that there is no motivation to combine references for teaching of claim 4

.In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention

Art Unit: 2162

where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of following the links until the Caribbean.htm is reached to Madnick's system in order to retrieve a relevant information corresponding to a user's request correctly and quickly.

c. Applicant argued that none cited art teaches claim 4 "following links contained within the web domain until the links have been exhausted or following the links until a predetermined limit is reached".

In response, Hennings teaches following the links until the Caribbean.htm is reached. Caribbean.html is represented as a predetermined limit (fig. 8).

For the above reasons, the cited references teach the claimed invention.

***Claim Rejections - 35 USC § 103***

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1, 2, 3, 5, 6, 10, 17-24, 26, 34, 37-38 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick in view of Iizuka et al (or hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982).

As to claim 1, Madnick teaches a method for extracting data from a network by a server (col. 3, lines 1-7; col. 1-2);

"enabling a database-structured query with at least one fundamental clause to be generated by a user" as the request translator receives a data request from data receiver 102 and translates the data request into a query at the wrapper generator 614. The converter query converts at least a portion of the query into a command to interact with a semi-structured data source such as HTML documents, flat files containing data that are not arranged as a relational database. The above information shows that a command is created based on the data request. The data request from data receiver 102 is represented as a user input. The command is represented as the database-structured query that is not generated by a user. The wrapper generator 614 is represented as a server (col. 2, lines 46-55, col. 2, lines 30-33);

“determining a web domain address on the network from which to extract the data” as determining a URL on the network to extract the data (table 2, col. 12, lines 1-10, lines 1-5);

“extracting the data from the web domain address directly by retrieving a non-database structured arrangement of data from the determined web domain address and performing the database-structured query upon the retrieved non-database structured arrangement of data” as at least a portion of the query is converted into one or more commands which can be used to interact with a semi-structured data source. Those commands are issued and data is extracted from the data source. In this case a source is located at an address or URL. The above information shows that the data is extracted from a semi-structured data source based on the address of the source and the command (fig. 7, col. 10, lines 25-32; col. 2, lines 2-8);

“providing extracted data from the determined web domain address in a data log directly to the user” as returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Madnick does not explicitly teach the claimed limitation “with at least one fundamental clause to be generated by a user; in data log directly to the user”.

Iizuka teaches the user interface unit receives a search request (query statement) consisting of search items and search condition (col. 13, lines 35-40).

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply lizuka's teaching user interface unit receives a search request (query statement) consisting of search items and search condition and Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user to Madnick's system in order to provide a result to a user quickly after retrieving data from a plurality of semi-structured document via open network without need a lot of time and labor to design and manage.

As to claim 2, Madnick teaches the claimed limitation "wherein creating the database-structured query further comprises, including a network address within the database-structured query indicating a starting point" as creating a command after converting at least a portion of a query, the command includes a network address as URL: <http://quotes.galt.com/>. This URL is indicated as a starting point (Table 2, col. 7, lines 25-32; col. 2, lines 5-10).

As to claim 3, Madnick teaches the claimed limitation "wherein the determined web domain address, includes at least one universal resource locator (URL)" as the URL (col. 12, lines 5-10, table 2).

As to claim 5, Madnick teaches the claimed limitation "wherein creating the database-structured query, further comprises, creating a regular expression within the

database-structured query used to determine the data to extract” creating regular expression with a specification file 706 as a command to determine the data to extract (col. 10, lines 2-5; col. 12, lines 5-10, table 2).

As to claim 6, Madnick teaches the claimed limitation “wherein directly extracting data from the web domain, further comprises, matching a plurality of patterns contained within the regular expression to the retrieved data to determine the data to extract” as each variable to be retrieved in a given state, the state description contains a pattern to be matched against the document or semi-structured data source. The above information shows that matching each pattern of each variable contained with the regular expression (col. 15, lines 1-10).

As to claim 10, Madnick teaches the claimed limitation “reshaping at least a portion of the extracted data for use by at least one data analysis software program” as extracted data is translated by the data translator from the data context of the data source into the data context associated with the initial request. It means that the extracted data is reshaped by translating. The above information shows that the system has included a data analysis software program to translate the extracted data (col. 3, lines 6-8).

As to claim 17, Madnick teaches the claimed limitations:



“a client computer system having a client network connection to the network and communicating with a server computer system” as (col. 3, lines 60-67; col. 4, lines 1-5);

“the server computer system having a server network connection to the network and communicating with the client computer system” as (col. 3, lines 60-67; col. 4, lines 1-5), “the server computer system further configured to perform actions, comprising:

receiving the database-structured query from the client computer system as the request translator receives a data request from data receiver 102 and translates the data request into a query. The converter query converts at least a portion of the query into a command to interact with a semi-structured data sources such as HTML documents, flat files containing data that are not arranged as a relational database. The above information shows that a command is created based on the data request. The data request from data receiver 102 is represented as a user input. The command is represented as the database-structured query (col. 2, lines 46-55, col. 2, lines 30-33);

“determining a web domain address on the network from which to extract at least a portion of the data relevant to the query, wherein the determined web domain address is provided by the database-structured query” (fig. 7, col. 10, lines 25-32; col. 2, lines 2-8);

“extracting at least the portion of the data from the web domain address directly by retrieving a non-database structured arrangement of data from the determined web domain address and performing the database-structured query upon the retrieved non-database structured arrangement of data” as (fig. 7, col. 10, lines 25-32; col. 2, lines 2-8);

“providing extracted data from the determined web domain address in a data log directly to the user” as returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Madnick does not explicitly teach the claimed limitation “the client creating a database-structured query with at least one fundamental clause, based, in part, on a user input; in data log directly to the user”.

lizuka teaches client 100 creating a database structured query based, in part, on a user input (fig. 39, col. 13, lines 35-45).

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply lizuka’s teaching of client 100 creating a database structured query based, in part, on a user input and Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user to Madnick’s system in order to provide a result to a user quickly after retrieving data from a plurality of semi-structured document via open network without need a lot of time and labor to design and manage.

As to claim 18, Madnick teaches the claimed limitation “wherein the database-structured query, further comprises, a network address within the database-structured query indicating a starting point” as (table 2, col. 12, lines 5-10).

As to claim 19, Madnick teaches the claimed limitation “a regular expression within the database-structured query used to determine the data to extract” as (col. 10, lines 2-5; col. 12, lines 5-10, table 2).

As to claim 20, Madnick teaches the claimed limitation “wherein the regular expression with the database-structured query further comprises at least one pattern used to determine the data to extract” as (col. 10, lines 2-5; col. 12, lines 5-10, table 2).

As to claim 21, Madnick teaches the claimed limitation “an editor for creating a template of regular expressions used to extract the data” as (col. 12, lines 5-10, table 2).

As to claim 22, Madnick teaches the claimed limitation “at least one data extraction engine to extract the data” as (col. 15, lines 25-35).

As to claim 23, Madnick teaches the claimed limitation “wherein the data extraction engine is a web crawler” as the wrapper generator 614 (col. 15, lines 25-35).

As to claim 24, Madnick teaches the claimed limitation “wherein the web domain address further comprises at least one link address for locating at least a portion of the data” as (col. 9, lines 55-67; col. 10, lines 1-5).

As to claim 26, Madnick teaches the claimed limitation "wherein the web domain address further comprises a link address, wherein at least another portion of the data is located with the link address" as (col. 9, lines 55-67; col. 10, lines 1-5).

As to claim 34, Madnick teaches the claimed limitation:

"generating a database structured query with at least one fundamental clause based, in part, on user input" as the request translator receives a data request from data receiver 102 and translates the data request into a query at the wrapper generator 614. The converter query converts a least a portion of the query into a command to interact with a semi-structured data sources such as HTML documents, flat files containing data that are not arranged as a relational database. The above information shows that a command is created based on the data request. The data request from data receiver 102 is represented as a user input. The command is represented as the database-structured query. The wrapper generator 614 is represented as a server (col. 2, lines 33-55, col. 8, lines 40-60);

"determining at least one webpage with the data, wherein the determination of the webpage is provided by the database-structured query" as extracting web pages that contains data by the commands (col. 9, lines 55-67; col. 10, lines 1-5);

"parsing the data at the at least one webpage in search of data that satisfies a query condition" as (col. 15, lines 1-10; table 2, col. 12, lines 1-20);

"wherein the data at the at least one web page is directly processed as though it is a searchable database" as the data receives 620 receives the web pages and

extracts the requested data from those pages. The above information shows that each web page or website is a searchable database (col. 10, lines 1-5);

“whereby a non-database structured arrangement of data is retrieved at the least one webpage and the database-structured query is performed upon the retrieved non-database structured arrangement of data” as each command is performed upon flat files containing data that are not arranged as a relational database at the website or web page (col. 2, lines 27-32; col. 9, lines 55-67; col. 10, lines 1-5);

“extracting at least a portion of the data from the retrieved non-database structured arrangement of data that satisfies the query condition” as extracting data at a web page that satisfies the query condition (col. 15, lines 1-20);

“providing extracted data from the determined web domain address in a data log directly to the user” as returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Madnick does not explicitly teach the claimed limitation “reshaping the extracted data to a predetermined format; in data log directly to the user”.

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

Iizuka teaches outputting the search result in a prescribed single format that is specific to each user. In particular, converting the search result into the item presentation styles of each user according to the style conversion data (col. 5, lines 5-10; col. 5, lines 35-40).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply lizuka's teaching of outputting the search result in a prescribed single format that is specific to each user. In particularly, converting the search result into the item presentation styles of each user according to the style conversion data lizuka's teaching of the apparatus returns directly the search result to a user via a user interface unit 11 to Madnick's system in order to provide a result to a user quickly after retrieving data from a plurality of semi-structured document via open network without need a lot of time and labor to design and manage, to retrieve data contained in a plurality of semi-structured documents over open network quickly, eliminate network traffic when server receives multiple user's request from at the same time and to provide a good view of a search result to a user's system for viewing easily.

As to claim 37, Madnick teaches the claimed limitation "wherein the structured query is generated to parse a limited portion of the data of the at least one webpage with the limits predetermined by the user" as (col. 12, lines 1-10, table 2).

As to claim 38, Madnick teaches the claimed limitation "wherein structured query is generated to search for at least one of a text string, a table, and a predefined list of words" as (col. 2, lines 30-55).

5. Claims 4, 35 and 36 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view of lizuka et al (or

hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Hennings et al (or hereinafter "Hennings") (US 6763496).

As to claim 4, Madnick does not explicitly teach the claimed limitation "following links contained within the web domain until the links have been exhausted or following the links until a predetermined limit is reached". Hennings teaches following the links until the Caribbean.htm is reached. Caribbean.html is represented as a predetermined limit (fig. 8).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of following the links until the Caribbean.htm is reached to Madnick's system in order to retrieve a relevant information corresponding to a user's request correctly and quickly.

As to claim 35, Madnick does not explicitly teach the claimed limitation "wherein the search of data is performed on at least a second webpage". Hennings teaches at least one link:<http://www.traveltickets.com> to <http://www.traveltickets.com/cruises> for locating Caribbean data to extract (fig. 8).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of at least one link:<http://www.traveltickets.com> to <http://www.traveltickets.com/cruises> for locating Caribbean data to extract to Madnick's system in order to retrieve a relevant information corresponding to a user's request correctly and quickly.

As to claim 36, Madnick does not explicitly teach the claimed limitation "wherein parsing the data of the at least one webpage further comprises following links included on the webpage and further parsing the data of webpages determined by the links included on the webpage". Hennings teaches a first web page comprises links and parsing data as shown in fig. 1B to determine links included on the web page (fig. 8).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of a first web page comprises links and parsing data as shown in fig. 1B to determine links included on the web page to Madnick's system in order to response to a customer's request for more detailed information about a document on a web page and further to retrieve a relevant information corresponding to a user's request correctly and quickly.

6. Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view of Iizuka et al (or hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Jammes.

As to claim 7, Madnick does not explicitly teach the claimed limitation "wherein creating the database structured query, further comprises, creating a condition expression with the database structured query describing how to scan the data at the determined web domain address for the data to extract". Jammes teaches as the following is one example of a name/value pair representing a query generated by the



Initial\_Event\_Handler to extract product data related to the root level group: query =select Product\_name, Product\_ID From Relationships, Groups where ID\_type = G and ID=1000 and relationship = Contains And (col. 22, lines 15-20).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Jammes's teaching of the following is one example of a name/value pair representing a query generated by the Initial\_Event\_Handler to extract product data related to the root level group: query =select Product\_name, Product\_ID From Relationships, Groups where ID\_type = G and ID=1000 and relationship = Contains to Madnick's system in order to retrieve data in different type of data structures corresponding to a user's request.

7. Claims 8-9 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view of Iizuka et al (or hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Jammes and Christianson et al (or hereinafter "Christianson") (US 6085186).

As to claim 8, Madnick discloses the claimed limitation subject matter in claim 1, except the claimed limitation "wherein directly extracting the data from the determined web domain, further comprises: retrieving data from the determined web domain address; reducing the retrieved data to a region of interest; and searching the region of interest for the data matching a predetermined regular expression".

Jammes teaches the claimed limitation “reducing the retrieved content to a region of interest” as an HTML coded result set: web/sedans.html>Sedans </A. This information shows the system reduced the retrieved content to a region of interest as Sedans (col. 22, lines 22-45).

Christianson teaches “searching the region of interest for the data matching a predetermined regular expression” as matching the returned html text with regular expression (col. 20, lines 65-67).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Jammes’s teaching of reducing the retrieved content to a region of interest as Sedans and Christianson teaching of matching the returned html text with regular expression to Madnick’s system in order to Madnick’s system in order to retrieve data in different type of data structures corresponding to a user’s request and to determine which sources are relevant to a given query, forwarding the query to the most relevant information sources, and further to provide regular expression component for creating modular hierarchical descriptions of regular expressions, for binding variables to the correct sub-strings recognized during pattern match to a response of an information source, for performing arbitrary action language statements with multiple variable bindings.

As to claim 9, Madnick discloses the claimed limitation subject matter in claim 1, except the claimed limitation “wherein directly extracting the data from the web domain, further comprises, storing the data matching the predetermined regular expression”.

Art Unit: 2162

Jammes teaches retrieving data records whose status fields match a predetermined status value and that a corresponding result set would be generated. This information shows that the system stores matched records (col. 26, lines 25-50).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Jammes's teaching of retrieving data records whose status fields match a predetermined status value and that a corresponding result set would be generated to Madnick's system in order to backup a system when the system is corrupted.

8. Claims 11-13, 15-16, 27-28, 30-33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view of Bates et al (or hereinafter "Bates") (US 6873982).

As to claim 11, Madnick teaches a computer-readable medium having computer-executable instructions for extracting data from a network (a memory having one or more commands to issue to the web page in order to retrieve the data from a network, col. 3, lines 21-26), "the computer-executable instruction enabling actions" (commands are enable for accessing the data and retrieving the data. Accessing and retrieving are represented as actions (col. 3, lines 20-26) comprises:

creating a database-structured query with at least one fundamental clause including a web domain address used for locating data, based, in part, on a user input"

Art Unit: 2162

as the request translator receives a data request from data receiver 102 and translates the data request into a query. The converter query converts at least a portion of the query into a command to interact with a semi-structured data source such as HTML documents, flat files containing data that are not arranged as a relational database. The above information shows that a command is created based on the data request. The data request from data receiver 102 is represented as a user input. The command is represented as the database-structured query (fig. 7; col. 2, lines 30-55; col. 8, lines 40-60);

“locating data based on the web domain address provided by the database-structured query” as the descriptor file 702 may be a directory of URL addresses which locate necessary information about the data source 104. The above information shows that the data source is located based on the URL addresses. The URL address is represented as the web domain address (col. 10, lines 27-30),

“extracting at least a portion of the located data directly by retrieving a non database structured arrangement of data from the located data and performing the database-structured query upon the retrieved non-database structured arrangement of data” as at least a portion of the query is converted into one or more commands which can be used to interact with a semi-structured data source. Those commands are issued and data is extracted from the data source. In this case a source is located at an address or URL. The above information shows that the data is extracted from a semi-structured data source based on the address of the source and the command (fig. 7, col. 10, lines 25-32; col. 2, lines 2-8);

“providing extracted data from the web domain address in a data log directly to the user” as returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Madnick does not explicitly teach the claimed limitation “ in data log directly to the user”.

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user to Madnick’s system in order to provide a result to a user quickly after retrieving data from a plurality of semi-structured document via open network without need a lot of time and labor to design and manage.

As to claim 12, Madnick teaches the claimed limitation “wherein the database-structured query, further comprises, a network address included within the database-structured query, further comprises, a network address included within the database-structured query indicating a starting point” as creating a command after converting at least a portion of a query, the command includes a network address as URL: <http://quotes.galt.com/>. Quotes.galt is indicated as a starting point (Table 2, col. 7, lines 25-32; col. 2, lines 5-10).

As to claim 13, "wherein the network address, further comprises at least one universal resource locator (URL)" as URL (col. 12, table 2).

As to claim 15, "wherein the database-structured query, further comprises, a regular expression within the database-structured query used to determine the data to extract" as a regular expression with the file 706 as the database-structured query (col. 12, table 2).

As to claim 16, "wherein the regular expression within the database-structured query, further comprises at least one pattern, used to determine the data to extract" as each variable to be retrieved in a given state, the state description contains a pattern to be matched against the document or semi-structured data source. The above information shows that matching each pattern of each variable contained with the regular expression (col. 15, lines 1-10).

As to claim 27, Madnick teaches the claimed limitations:

"creating a database-structured query with at least one fundamental clause at the server based, in part, on a user input" as the request translator receives a data request from data receiver 102 and translates the data request into a query at the wrapper generator 614. The converter query converts a least a portion of the query into a command to interact with a semi-structured data sources such as HTML documents, flat

files containing data that are not arranged as a relational database. The above information shows that a command is created based on the data request. The data request from data receiver 102 is represented as a user input. The command is represented as the database-structured query. The wrapper generator 614 is represented as a server (fig. 7, col. 2, lines 30-55; col. 8, lines 40-60);

“determining a website to search based in part on the database-structured query” as determining a URL on the network to extract the data implies determines a website (table 2, col. 12, lines 1-10, lines 1-5);

“extracting at least a portion of the data relevant to the database-structured query at the website directly based on the database-structured query” as extracting the requested web pages to the wrapper generator 614 in response to the transmitted commands (col. 9, lines 55-67; col. 10, lines 1-5);

“wherein the website is processed as a searchable database” as the data receives 620 receives the web pages and extracts the requested data from those pages. The above information shows that each web page or website is a searchable database (col. 10, lines 1-5);

“whereby a non-database arrangement of data is retrieved from the website and the database-structured query is performed upon at least the retrieved non-database arrangement of the data ” as each command is performed upon flat files containing data that are not arranged as a relational database at the website or web page (col. 2, lines 27-32; col. 9, lines 55-67; col. 10, lines 1-5).

“to extract at least the portion of the data from the retrieved non-database arrangement of the data” as extracting the data from the HTML documents that the non-database arrangement of the data (col. 2, lines 27-32; col. 10, lines 1-5).

“providing extracted data from the website in a data log directly to the user” as returning extracted data from address in source not in data log indirectly to the user (col. 9, lines 49-63, fig. 6).

Madnick does not explicitly teach the claimed limitation “in data log directly to the user”.

Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user (fig. 10, col. 11, lines 60-67; col. 12, lines 1-27).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Bates teaches storing extracted documents as results in result cache and returning a first document as a first result from the result cache directly to a user to Madnick’s system in order to provide a result to a user quickly after retrieving data from a plurality of semi-structured document via open network without need a lot of time and labor to design and manage.

As to claim 28, Madnick teaches the claimed limitation “parsing the database-structure query to determine at least one link to search at the website” as (col. 12, lines 1-20, table 2).



As to claim 30, Madnick teaches the claimed limitation “determining what data to extract based in part on the database-structured query and the provided web domain address” as (col. 12, lines 1-20, table 2).

As to claim 31, Madnick and Iizuka teach the claimed limitation subject matter in claim 27, Iizuka further teaches the claimed limitation “wherein extracting data based in part on at least one of an Hypertext Markup Language (HTML) table, a binary file, and a matching pattern” as extracting data based on an HTML table (col. 14, lines 34-40).

As to claim 32, Madnick teaches the claimed limitation “reshaping the extracted data for at least one of a database, a spreadsheet, Extensible Markup Language (XML) display, and a statistical tool” as (col. 3, lines 1-8).

As to claim 33, Madnick teaches the claimed limitation “wherein the website is a starting website based in part on the database-structured query” as (col. 10, lines 1-5).

9. Claims 14, 25 and 29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter “Madnick”) (US 5913214) in view of Bates and further in view of Hennings et al (or hereinafter “Hennings”) (US 6763496).

As to claim 14, Madnick does not explicitly teach the claimed limitation “at least one link to another web domain address for locating data to extract”.

Art Unit: 2162

Hennings teaches at least one link:<http://www.traveltickets.com> to

<http://www.traveltickets.com/cruises> for locating Caribbean data to extract (fig. 8).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of at least one link:<http://www.traveltickets.com> to <http://www.traveltickets.com/cruises> for locating Caribbean data to extract to Madnick's system in order to response to a customer's request for more detailed information about a document on a web page.

As to claim 25, Madnick does not explicitly teach the claimed limitation "at least one link address that is followed to locate data to extract until a predetermined number of links is reached". Hennings teaches following the links until the Caribbean.htm is reached. Caribbean.html is represented as a predetermined limit (fig. 8).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Hennings' s teaching of following the links until the Caribbean.htm is reached to Madnick's system in order to response to a customer's request for more detailed information about a document on a web page.

As to claim 29, Madnick teaches the claimed limitation "determining at least one other website to search based in part on the database-structured query and a provided web domain address" as (col. 9, lines 55-67; col. 10, lines 1-5).

Madnick does not explicitly teach the claimed limitation "extracting at least another portion of the data at the at least one other website based on the database-

Art Unit: 2162

structured query and the provided web domain address, wherein the at least one other website include a non-database structured arrangement of data that is processed as a searchable database". Hennings teaches extracting Golfing data at a second web page. This web page includes a HTML document as a non-database structured arrangement of data (fig. 8).

It would have been obvious to a person of an ordinary in the art at the time the invention was made to apply Hennings's teaching of extracting Golfing data at a second web page to Madnick's system in order to retrieve data contained in a plurality of semi-structured documents over a network.

10. Claim 42 is rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view Iizuka et al (or hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Fleskes (US 6529910).

As to claim 42, Madnick does not explicitly teach the claimed limitation "providing authentication data to the web domain".

Fleskes teaches providing authentication data to a domain (col. 2, lines 1-10).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Fleskes's teaching of providing authentication data to a domain to Madnick's system in order to restrict access for modify web page without permission and provide a user sufficient security access rights.

11. Claim 41 is rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view lizuka et al (or hereinafter "lizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Rheume (US 6247018) .

As to claim 41, Madnick does not explicitly teach the claimed limitation "wherein the at least one fundamental clause includes a request to parse an HTML table, and wherein extracting the data further comprise extracting data from HTML table".

Rheume teaches parsing an HTML table and extracting the data from HTML table (col. 11, lines 10-15, figs. 8A-8B).

It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Rheume's teaching of parsing an HTML table and extracting the data from HTML table to Madnick's system in order to to help a user to search/retrieve/store a portion of a document easily and quickly in large database and further retrieve a HTML page or a group of related HTML pages in an HTML frameset from a user specified URL or from a disk file.

12. Claim 43 is rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view lizuka et al (or hereinafter "lizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Eckes (US 6243832).

As to claim 43, Madnick does not explicitly teach the claimed limitation "wherein the extracted data includes at least one binary file".

Eckes teaches loading binary file (col. 2, lines 40-45).

It would have been obvious to a person of an ordinary skill in the art at the time invention was made to apply Eckes's teaching of downloading binary file to Madnick's system in order to allow faster retrievals and reduced resource consumption requirements.

13. Claims 41 and 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over Madnick et al (or hereinafter "Madnick") (US 5913214) in view of Iizuka et al (or hereinafter "Iizuka") (US 6424980) and Bates et al (or hereinafter "Bates") (US 6873982) and further in view of Jammes

As to claim 41, Madnick does not explicitly teach the claimed limitation "wherein the at least one fundamental clause includes a request to parse an HTML table, and wherein extracting the data further comprise extracting data from HTML table".

Jammes teaches parsing HTML file and extracting data from HTML file (fig. 18). It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Jammes's teaching of parsing HTML file and extracting data from HTML file to Madnick's system in order to help a user to search/retrieve/store a portion of a document easily and quickly in large database and further retrieve a

Art Unit: 2162

HTML page or a group of related HTML pages in an HTML frameset from a user specified URL or from a disk file.

As to claim 44, Madnick teaches the claimed limitation "wherein the server computer is further configured to perform the actions including: providing a stored database-structure query to the client computer system upon user input request" as (col. 2, lines 33-55, col. 8, lines 40-60).

Madnick does not explicitly teach "storing the database-structured query".

Jammes teaches storing SQL queries in HTML template file (col. 9, lines 10-20). It would have been obvious to a person of an ordinary skill in the art at the time the invention was made to apply Jammes's teaching of storing SQL queries to Madnick's system in order to help a user to search/retrieve/store a portion of a document easily and quickly in large database.

### ***Conclusion***

14. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Kraft (US 6633867) .

**Contact Information**

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Cam Y T. Truong whose telephone number is (571) 272-4042. The examiner can normally be reached on Monday to Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Breene can be reached on (571) 272-4107. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Cam Y T Truong  
Primary Examiner  
Art Unit 2162